

Bayesian estimation of Hyperparameters in MRI through the Maximum Evidence Method

Damián E. Oliva¹, Roberto A. Isoardi² and Germán Mato³

¹*Facultad de Ciencias Exactas, Universidad de Buenos Aires, Argentina*

²*Escuela de Medicina Nuclear, Mendoza, Argentina*

³*Grupo de Física Estadística, Centro Atómico Bariloche, Argentina*

E-mail: risoardi@ieee.org

Abstract

Bayesian inference methods are commonly applied to the classification of brain Magnetic Resonance images (MRI). We use the Maximum Evidence (ME) approach to estimate the most probable parameters and hyperparameters for models that take into account discrete classes (DM) and models accounting for the partial volume effect (PVM). An approximate algorithm was developed for model optimization, since the exact image inference calculation is computationally expensive. The method was validated using simulated images and a digital phantom. We show that the Evidence is a very useful figure for error prediction, which is to be maximized respect to the hyperparameters. Additionally, it provides a tool to determine the most probable model given measured data.

1. Introduction

A wide variety of techniques are available for statistical inference of magnetic resonance images of the brain. An extensive review can be found in the literature [1]. Another set of useful algorithms are based on anisotropic filtering [2-5].

In this work we use unsupervised algorithms to infer tissue distribution in brain MRI [1]. Such unsupervised algorithms do not require previous training for tissue classification, since the user introduces the known image data in the *a priori* model. One example is the discrete algorithm [6], which accounts for magnetic field heterogeneities [7] and includes an interaction potential between pixels in the *a priori* distribution [8,9].

The main motivation is to develop an automatic

algorithm that is able to determine the typical spatial scales that appear in the image. In this way, we are able to discriminate structures as small as possible, but without fitting spurious signals generated by noise.

This objective is achieved by introducing two new features that were added to the discrete algorithm: partial volume effect correction and the approximate estimation of the pixel interaction factor. We chose the Maximum Evidence criterion (ME) [10], as the Bayesian method for model optimization and comparison.

Model optimization usually involves the minimization of an energy function, which can be done by Simulated Annealing [8,9]. To compare different models, it is often necessary to solve complex integrals using Monte Carlo tools [11]. Both procedures are so computationally intensive that they become impractical for medical applications. It is therefore attractive to develop approximate Bayesian methods, capable of reducing calculation time considerably. This paper focuses on the calculation of the approximate Evidence for the hyperparameters (parameters that control the behavior of many pixels) and for the different models. We show that Evidence is a very useful figure for error prediction.

The nomenclature and notation used throughout the work are introduced in Section 2. Section 3 summarizes the Maximum Evidence method [10,11], used for the estimation of optimal parameters and hyperparameters. This procedure is carried out in three steps. First, parameters are optimized for each pixel. Second, optimization is done for the hyperparameters. In the last step, the different models are compared. In Section 4, the same steps are implemented for the discrete model, whereupon the Zhang algorithm [6] is obtained, including the approximate Bayesian estimation for the pixel interaction factor. Section 5

describes the same procedure for the partial volume model. In Section 6 the algorithm is tested on simulated images. Finally, its validation is assessed using a digital phantom [12].

2. Nomenclature and Notation

1) *Models*: The different models are denoted as H_i .

H_1 : Discrete Model (DM), H_2 : Partial Volume Model (PVM)

2) *Data*: vector \mathbf{d} represents the image, typically $N=256 \times 256$

3) *Parameters and hyperparameters*: The parameter vector of model H_i is denoted by \mathbf{c} . Each component \mathbf{c}_i ($i = 1, \dots, N$), belongs to one of L tissue classes, associated with gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and background (BKG). The hyperparameter of H_i is: $\gamma_i = (\mu, \sigma, \beta, L)$. Where μ and σ are the mean intensity and standard deviation for each class ($1, \dots, L$) and β is the pixel interaction factor.

The parameter vector of model H_2 is denoted by \mathbf{a} , i.e. \mathbf{a}_i is the fraction of the most probable class for pixel i . Likewise, the hyperparameter of H_2 is: $\gamma_2 = (\mu, \sigma, \beta, L)$.

3. Maximum Evidence Criterion

The Bayesian inference is formulated in three steps: *1st Step*) Search for the most probable parameters *2nd Step*) Search for the most probable hyperparameters and *3rd Step*) Determination of the most probable model [10].

3.1 Most probable parameters (1st Step)

Given measured data \mathbf{d} and model H_i , which are the most probable parameters \mathbf{w} (where \mathbf{w} corresponds to \mathbf{c} or \mathbf{a} according to the model) in the model? According to Bayes' rule:

$$P(\mathbf{w}|\mathbf{d}, \gamma_i, H_i) = \frac{P(\mathbf{d}|\mathbf{w}, \gamma_i, H_i) \cdot P(\mathbf{w}|\gamma_i, H_i)}{P(\mathbf{d}|\gamma_i, H_i)} \quad (1)$$

where $P(\mathbf{d}|\mathbf{w}, \gamma_i, H_i)$ is the *Likelihood* distribution (noise model), $P(\mathbf{w}|\gamma_i, H_i)$ is the *a priori* distribution and $P(\mathbf{d}|\gamma_i, H_i)$ is the *Evidence* of hyperparameter γ_i . The distribution in the denominator of Eq. (1) can be written as:

$$P(\mathbf{d}|\gamma_i, H_i) = \int P(\mathbf{d}, \mathbf{w}|\gamma_i, H_i) \cdot d\mathbf{w} = \int P(\mathbf{d}|\mathbf{w}, \gamma_i, H_i) \cdot P(\mathbf{w}|\gamma_i, H_i) d\mathbf{w} \quad (2)$$

If \mathbf{w} is discrete, the integral in Eq. (2) becomes a summation over all possible values of \mathbf{w} . Since

$P(\mathbf{d}|\gamma_i, H_i)$ does not depend on \mathbf{w} ,

$$P(\mathbf{w}|\mathbf{d}, \gamma_i, H_i) \propto P(\mathbf{d}|\mathbf{w}, \gamma_i, H_i) \cdot P(\mathbf{w}|\gamma_i, H_i) \quad (3)$$

The most probable parameter \mathbf{w}_{MP} in the confidence interval σ_{MP} (error bars) can be found by maximizing the posterior distribution with respect to \mathbf{w} . This calculation is the first step of the Bayesian inference. As the posterior distribution is usually a very complex function, the absolute maximum can be found using Simulated Annealing and Monte Carlo techniques [8,9]. However, such methods are computationally expensive, and a faster alternative is provided by the Quadratic Approximation [10], the Gradient Descent method for continuous variable models [5] or approximate techniques such as Iterated Conditional Modes for the discrete case [13].

3.2 Most probable hyperparameters (2nd Step)

Given a set of hyperparameters γ_i for model H_i and measured data \mathbf{d} , which is the most probable γ_i ? The posterior probability of hyperparameter γ_i can be found using Bayes' rule:

$$P(\gamma_i|\mathbf{d}, H_i) \propto P(\mathbf{d}|\gamma_i, H_i) \cdot P(\gamma_i|H_i) \quad (4)$$

Assuming that all hyperparameters are equally probable *a priori*,

$$P(\gamma_i|H_i) = \text{constant} \quad \forall \gamma_i \quad (5)$$

It follows that the posterior probability of each hyperparameter is proportional to its evidence:

$$P(\gamma_i|\mathbf{d}, H_i) \propto P(\mathbf{d}|\gamma_i, H_i) \quad (6)$$

Under the above assumptions, *the most probable hyperparameter is the one with the maximum evidence*. As described in [10], (6) can be approximated by:

$$P(\mathbf{d}|\gamma_i, H_i) \propto P(\mathbf{d}|\mathbf{w}_{MP}, \gamma_i, H_i) \cdot P(\mathbf{w}_{MP}|\gamma_i, H_i) \\ \text{Log}[P(\gamma_i|\mathbf{d}, H_i)] \propto -\frac{(\mathbf{d} - \mathbf{f}(\mathbf{w}_{MP}))^2}{2\sigma^2} + \text{Log}\left[\frac{\sigma_{w|d}}{\Delta \mathbf{w}^0}\right] \quad (7)$$

where $OF = \sigma_{w|d} / \Delta \mathbf{w}^0$ is the Occam factor, $\Delta \mathbf{w}^0$ is the volume of the *a priori* distribution in parameter space

and \mathbf{f} is some fitting function. As an example, suppose that our model H_i fits data \mathbf{d} with a γ_i -degree polynomial \mathbf{f} with coefficients \mathbf{w} . Note that $\text{Log}[\text{Evidence}]$ can behave in two different ways with increasing γ_i . $\text{Log}[\text{Evidence}]$ increases as the term $(\mathbf{d} - \mathbf{f}(\mathbf{w}_{\text{MP}}))^2$ is smaller, but on the other hand, it decreases due to a larger $\Delta^0 \mathbf{w}$ and decreasing OF . In this way, the image inference models can be described as algorithms which fit surfaces of varying flexibility (controlled by hyperparameter γ_i). The most probable γ_i can be found by maximizing (6).

In general, our current models have many degrees of freedom and the integration in (2) is performed over L^N possible class configurations. One possibility is to evaluate such integration by Monte Carlo method. Our approach, as described in the following sections, is to approximate the Evidence. In this way the number of configurations to evaluate is N^*L .

3.3 Most probable model (3rd Step)

Given the set of models H_1, \dots, H_n for measured data \mathbf{d} . Which one is most probable? From the previous section, the Maximum Evidence model is the most probable one:

$$P(\mathbf{d}|H_i) \propto \int d\gamma_i P(\mathbf{d}, \gamma_i | H_i) \approx P(\mathbf{d}|\hat{\gamma}_i, H_i) \cdot \Delta \hat{\gamma}_i \cdot P(\hat{\gamma}_i | H_i) \quad (8)$$

with $P(\gamma_i | H_i) = 1/\Delta^0 \gamma_i$ where $\Delta^0 \gamma_i$ is the volume of the *a priori* distribution for hyperparameter γ_i .

4. The Maximum Evidence Method for the Discrete Model

The Evidence of a model can be increased by reducing its flexibility. For that reason, we shall first consider the discrete model, since it has few degrees of freedom. Next, the complexity of the model will be increased in order to reduce the quadratic error (correction of partial volume).

4.1 Discrete Model Assumptions

- The mean intensity μ_j for each tissue (class j) is well defined.
- Each pixel belongs to only one of four possible classes c_j : BKG, CSF, WM and GM.
- Gaussian noise distribution [14].
- In most cases, each pixel belongs to the same class as its neighbors.

From the above assumptions, the probability distribution for pixel i can be written as:

$$P(d_i | c_i, \mu_{c_i}, \sigma_{c_i}, H_1) = \frac{1}{\sqrt{2\pi\sigma_{c_i}^2}} \exp\left(-\frac{(d_i - \mu_{c_i})^2}{2\sigma_{c_i}^2}\right) \quad (9)$$

Moreover, if the noise is uncorrelated:

$$P(\mathbf{d}|\mathbf{c}, \boldsymbol{\mu}, \boldsymbol{\sigma}, H_1) = \prod_{i=1}^N P(d_i | c_i, \mu_{c_i}, \sigma_{c_i}, H_1) \quad (10)$$

4.2 "A priori" distribution

With hypothesis d), image pixels in the same neighborhood are 'a priori' expected to have similar intensity values. This property allows the MR image to be described as a MRF (Markov Random Field) [8-10] where each pixel c_i is related to K neighbors c_m in neighborhood $N(i)$ through an interaction potential:

$$V_1(c_i, c_m) = \sum_{c_m \in N(i)} \delta(c_i - c_m) = N_{\text{diff}} \quad (11)$$

where N_{diff} is the number of pixels of $N(i)$ with a class other than c_i

$$P(\mathbf{c}|\boldsymbol{\beta}, H_1) = \frac{\exp(-\beta V_1(\mathbf{c}))}{Z_c(\boldsymbol{\beta})} \quad (12)$$

$$Z_c(\boldsymbol{\beta}) = \sum_{\forall \mathbf{c}} \exp(-\beta V_1(\mathbf{c}))$$

and $Z_c(\boldsymbol{\beta})$ is the partition function (or normalization constant) which depends on hyperparameter $\boldsymbol{\beta}$ (pixel interaction factor).

4.3 Bayesian inference for the Discrete Model (1st step)

Approximating the joint distribution (Appendix A) and setting $\mathbf{w}=\mathbf{c}$, we obtain the approximate posterior distribution:

$$\text{Log}[P(\mathbf{c}|\mathbf{d}, \gamma_1, H_1)] = -\sum_1^N \left\{ \begin{array}{l} e(c_i | \hat{c}_{N(i)}) - \text{Log}[Z^{(i)}(\gamma_1)] \\ - \text{Log}[P(d_i | \gamma_1, H_1)] \end{array} \right\}$$

$$e(c_i | \hat{c}_{N(i)}) = \frac{(d_i - \mu_{c_i})^2}{2\sigma_{c_i}^2} + \frac{\beta}{2} V_1(c_i | \hat{c}_{N(i)}) \quad (13)$$

In this way, maximizing the posterior probability of class \mathbf{c} is equivalent to minimizing $e(c_i, c_m)$ for each pixel, evaluating its neighbors for their most probable class.

4.4 Bayesian inference for the Discrete Model (2nd step)

As shown in (2), the exact calculation of the evidence requires the addition of L^N terms. Using A.4 we approximate $\text{Log}[\text{Evidence}]$ for the whole image as the sum of the evidence for each pixel (with neighboring pixels $c_{N(i)}$ in their most probable configuration). With this approach, the number of terms needed drops to $N \cdot L$. Using $\mathbf{w}=\mathbf{c}$:

$$\begin{aligned} \text{Log} [P(\mathbf{d}|\hat{c}_{N(i)}, \gamma_1, H_1)] &\approx \sum_{i=1}^N \text{Log} [P(d_i|\hat{c}_{N(i)}, \gamma_1, H_1)] = \\ &= \sum_{i=1}^N \text{Log} \left[\sum_{c_i=1}^L P(d_i, c_i|\hat{c}_{N(i)}, \gamma_1, H_1) \right] \end{aligned} \quad (14)$$

with the joint distribution for pixel i :

$$P(d_i, c_i|\hat{c}_{N(i)}, \gamma_1, H_1) = \frac{\exp\left(-\frac{(d_i - \mu_{c_i})^2}{2\sigma_{c_i}^2}\right) \exp\left(-\frac{\beta}{2} V_1(c_i|\hat{c}_{N(i)})\right)}{\sqrt{2\pi\sigma_{c_i}^2} Z_c(\beta)} \quad (15)$$

To obtain the most probable hyperparameter γ_1 , we derive the approximate evidence:

$$\begin{aligned} \frac{\partial}{\partial \gamma_1} \left[\sum_i \text{Log} \left(\sum_{c_i} P(d_i, c_i|\hat{c}_{N(i)}, \gamma_1) \right) \right] &= \\ = \sum_i \left(\frac{1}{\sum_{c_i} P(d_i, c_i|\hat{c}_{N(i)}, \gamma_1)} \sum_{c_i} \frac{\partial}{\partial \gamma_1} P(d_i, c_i|\hat{c}_{N(i)}, \gamma_1) \right) &= 0 \end{aligned} \quad (16)$$

The solution of (16) yields a set of equations for the most probable hyperparameters:

$$\begin{aligned} \hat{\mu}_j &= \frac{1}{N} \sum_i P(c_i = j|\hat{c}_{N(i)}, d_i, \gamma_1) \cdot d_i \\ (\hat{\sigma}_j)^2 &= \frac{1}{N} \sum_i P(c_i = j|\hat{c}_{N(i)}, d_i, \gamma_1) \cdot (d_i - \hat{\mu}_j)^2 \end{aligned} \quad (17)$$

For the discrete model case, the most probable hyperparameter β can be found analytically:

$$\beta = 2 \cdot \text{Log} \left[\left(\frac{N \cdot K}{\sum_{i=1}^N \sum_{j=1}^L \sum_{m=1}^K P(c_i = j|\hat{c}_{N(i)}, d_i, \gamma_1) V_1(c_i = j|\hat{c}_m)} - 1 \right) (L-1) \right] \quad (18)$$

where K is the number of neighbors of a given pixel or pixel and:

$$P(c_i = j|\hat{c}_{N(i)}, d_i, \gamma_1) = \frac{P(d_i|c_i = j, \gamma_1) \cdot P(c_i = j|\hat{c}_{N(i)}, \gamma_1)}{\sum_{c_i=1}^L P(d_i, c_i|\hat{c}_{N(i)}, \gamma_1)} \quad (19)$$

4.5 Algorithm for the Discrete Model

The discrete algorithm, obtained by the Maximum Evidence approach, turns out quite similar to that of Zhang [6], except we now determine the most probable interaction factor β using (18), after having estimated μ and σ .

5. Partial Volume Model

As stated in section IV, one of the assumptions of the Discrete Model is that there is only one tissue per pixel. This is obviously not valid for those pixels with more than one tissue. Therefore, in the next approach two classes are allowed for each pixel [15,16].

In the new scheme, the parameter $\mathbf{w}=\mathbf{a}$ is a $N \times 2$ matrix where $[\mathbf{a}]_{i,1} = a_{i,a}$ and $[\mathbf{a}]_{i,2} = a_{i,b}$ ($0 < a_{i,\cdot} < 1$) are the fractions of most probable tissue and second most probable tissue for pixel i , respectively. Assuming the noise as Gaussian and independent, we write the Likelihood of the PVM model:

$$\begin{aligned} P(d_i|a_{i,a}, a_{i,b} = 1 - a_{i,a}, \gamma_2, H_2) &= \\ = \frac{\exp\left(-\frac{(d_i - (a_{i,a} \cdot \mu_a + a_{i,b} \cdot \mu_b))^2}{2(a_{i,a}^2 \cdot \sigma_a^2 + a_{i,b}^2 \cdot \sigma_b^2)}\right)}{\sqrt{2\pi(a_{i,a}^2 \cdot \sigma_a^2 + a_{i,b}^2 \cdot \sigma_b^2)}} \end{aligned} \quad (20)$$

As in (10), the noise is assumed not correlated.

5.1 'A priori' distribution for the PVM

The 'A priori' distribution for the PVM is:

$$P(\mathbf{a}|\beta, H_2) = \frac{\exp(-\beta V_2(\mathbf{a}))}{Z_c(\beta)} \quad (21)$$

Adopting a square potential (associated with a low-pass filter):

$$V_2(\mathbf{a}) = \sum_{i=1}^N \sum_{m \in N(i)} |\mathbf{a}_i - \mathbf{a}_m|^2 = \sum_{i=1}^N \sum_{m \in N(i)} 2(a_{i,a} - a_{m,a})^2 \quad (22)$$

In this way, the PVM can be considered as an extension of the Discrete Model.

5.2 Bayesian inference for the PVM (1st step)

To simplify the algorithm, it is assumed that tissue fractions for pixel i can take the N_p values: $a_{i,a}=(1/N_p, \dots, (N_p-1)/N_p, 1)$.

From approximation A.4, the number of evaluations increases linearly with N_p . The approximate energy to be minimized in this case will be:

$$E(\mathbf{a}) = \sum_{i=1}^N e(\mathbf{a}_i | \hat{\mathbf{a}}_{N(i)}) \quad (23)$$

$$e(\mathbf{a}_i | \hat{\mathbf{a}}_{N(i)}) = e_D(\mathbf{a}_i) + \frac{\beta}{2} \cdot V_2(\mathbf{a}_i | \hat{\mathbf{a}}_{N(i)})$$

In this way we find the $a_{i,a}$ that minimizes the energy for each pixel with its neighbors in the most probable configuration. This procedure is repeated for all image pixels.

5.3 Bayesian inference for the PVM (2nd step)

Using the Evidence approximation (see appendix A) and setting $\mathbf{w}=\mathbf{a}$:

$$\begin{aligned} \text{Log}[P(\mathbf{d} | \gamma_2, H_2)] &\approx \sum_{i=1}^N \text{Log}[P(d_i | \hat{a}_{N(i),a}, \gamma_2, H_2)] = \\ &= \sum_{i=1}^N \text{Log} \left[\sum_{k=1}^{N_p} P \left(d_i, a_{i,a} = \frac{k}{N_p} \mid \hat{a}_{N(i),a}, \gamma_2, H_2 \right) \right] \end{aligned} \quad (24)$$

$$P(d_i, a_{i,a} | \hat{a}_{N(i),a}, \gamma_2, H_2) = \frac{\exp(-e_D(a_{i,a}))}{Z_D^{(i)}(\sigma, a_{i,a})} \cdot \frac{\exp\left(-\beta \frac{V_2}{2}(a_{i,a} | \hat{a}_{N(i),a})\right)}{Z_c^{(i)}(\beta)} \quad (25)$$

where $Z_c^{(i)}$ is:

$$\begin{aligned} Z_c^{(i)}(\beta) &= \prod_{i=1}^K Z_m^{(i)}(\beta) \\ Z_m^{(i)}(\beta) &= \sum_{k=0}^{N_p} \exp\left(-\beta \frac{V_2}{2} \left(a_{i,a} = \frac{k}{N_p} \mid \hat{a}_{N(i),a} \right) \right) \end{aligned} \quad (26)$$

The optimal values for β and N_p can be selected by maximizing $\text{Log}[\text{Evidence}]$ (24).

5.4 Algorithm description

As previously stated, the Partial Volume Model can be thought of as an extension of the Discrete Model. To optimize the PVM, the energy E will be minimized relative to the DM and then for the PVM. The

procedure is as follows:

- 1) Classify using the discrete algorithm.
- 2) Determine the partial volume pixels (PV). To do that, we find the posterior probability that pixel i belongs to class j (19). We define PV pixels as those with $P(\mathbf{d} | \mathbf{w}, \gamma_b, H_1) < 0.95$. PV pixels are mostly found in tissue transition regions.
- 3) Use (19) to determine the two most probable classes for PV pixels.
- 4) Take $a_{i,a}=(1/N_p, 2/N_p, \dots, (N_p-1)/N_p, 1)$ and minimize the energy $e(a_{i,a})$ in (23) for each case. This is done across all PV pixels.
- 5) If $e(a_{i,a})$ decreases, the pixels in the neighborhood of i are considered as PV pixels.
- 6) Calculate the evidence (14). Go back to step 4 until the Evidence converges to a local maximum.

The optimizations of models DM and PVM only differ in PV pixels. However, that distinction is not really necessary and it was done only to reduce the Occam Factor and to save calculation time (thus all image pixels can be defined as PV pixels). The optimal values for β and N_p can be selected by maximizing the approximate $\text{Log}[\text{Evidence}]$ (24).

6. Algorithm Validation

6.1 PVM validation using simulated 1D images

As previously indicated, the PVM introduces parameter \mathbf{a} to model the partial volume effect. Also, the $a_{i,a}$ were discretized into N_p possible values. This explains why *a priori* each pixel has more flexibility (more degrees of freedom) than for the DM. For this reason, the Occam Factor is expected to penalize the PVM relative to the DM. Furthermore, the interaction factors (β) reduce the flexibility of both models.

The problem can be formulated by the following questions: which is the most probable β for the PVM? and also, is the PVM more probable than the DM? These two questions correspond to steps 2 and 3 of the Bayesian inference (Section III).

First, the problem will be analyzed for a 1-D, simulated image:

$$\begin{aligned} f(x_i) &= A \cdot \tanh(\text{sat} \cdot x_i) + v \\ v &: N(0, \sigma_R) \quad \text{Gaussian noise.} \end{aligned} \quad (27)$$

This image is chosen because is one of the simplest

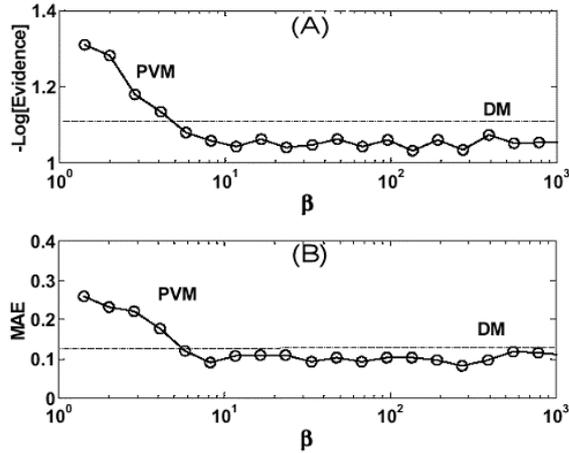


Figure 1: Comparison between the DM and PVM with a simulated 1D image (see Eq. 27) with $A=1$, $\sigma=0.7$, $\text{sat}=0.3$, $N=64$, $K=2$. (A) $-\text{Log}[\text{Evidence}]$ vs. β (B) MAE vs. β . Dotted lines correspond to the inference of the DM using (18). The curves with "o" correspond to $-\text{Log}[\text{Evidence}]$ and MAE of PVM as a function of β . The difference between the dotted line and the "o" curve is proportional to $1/\text{sat}$.

cases with a non-trivial transition between different intensities. The width of the transition between different regions is regulated by the parameter sat . The noise is also introduced in the simplest possible way: independent pixel to pixel and with a Gaussian distribution.

Fig. 1.A shows $-\text{Log}(\text{Evidence})$ from (24) and Fig. 1.B shows the Mean Absolute Error (MAE) for the DM and the PVM. For the DM, β is chosen according to (18). For the PVM, $-\text{Log}(\text{Evidence})$ and the MAE are shown as a function of β .

For small β , the DM shows a better performance than the PVM, due to the penalization of the latter by the OF. This behavior can be observed from either the Evidence analysis or the MAE. This example shows that the Maximum Evidence Criterion (MEC) is satisfactory for the inference of the most probable value of β .

Variable sat (27) controls the width of the partial volume region, hence the PVM is expected to perform better than the DM when sat is decreased. This behavior was found for both the MAE and the Evidence.

6.2 Digital Phantom

The validation process seeks to verify if the algorithm complies with the design requirements for real world

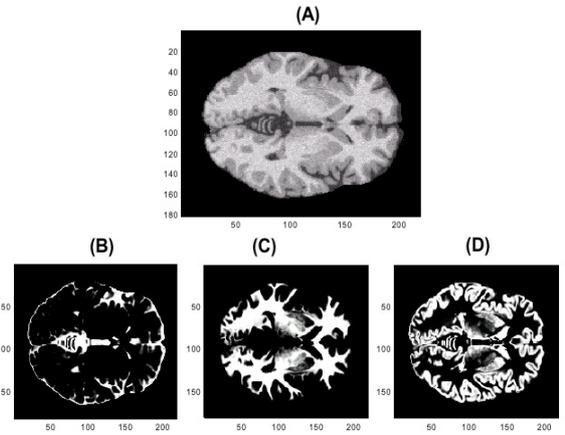


Figure 2: A) Transaxial slice of a MR(T1) brain digital phantom [12]. B), C), D): Fraction of CSF, WM and GM tissues, respectively.

images. The main difficulty in testing algorithm performance lies in the fact that the actual distribution of "in vivo" tissues is unknown. Realistic digital phantoms such as the one developed at the McConnell Brain Image Processing Centre [12], provide a convenient alternative to that problem. It includes a Magnetic Resonance database of the tissue fractions, which constitute the basis to simulate real data acquisition: noise and magnetic field heterogeneities. For this study a T1-weighted volume was generated ($181 \times 217 \times 181$), the selected amount of noise was 5% and the slice thickness was 1 mm, without simulation of magnetic field heterogeneities.

Since this work is intended for brain tissue segmentation, all extra-cerebral structures were previously removed with a mask, so that the tissues of interest are: GM, WM, CSF and BKG. The fraction of tissue j in pixel i is $t_{i,j}$. Fig. 2A shows the simulated image. Tissue fractions are shown in Fig. 2.B,C,D.

6.3 Classification error

The classification error can be measured as:

$$MAE = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^L p(j) \cdot \text{Abs}(p_{i,j} - t_{i,j}) \quad (28)$$

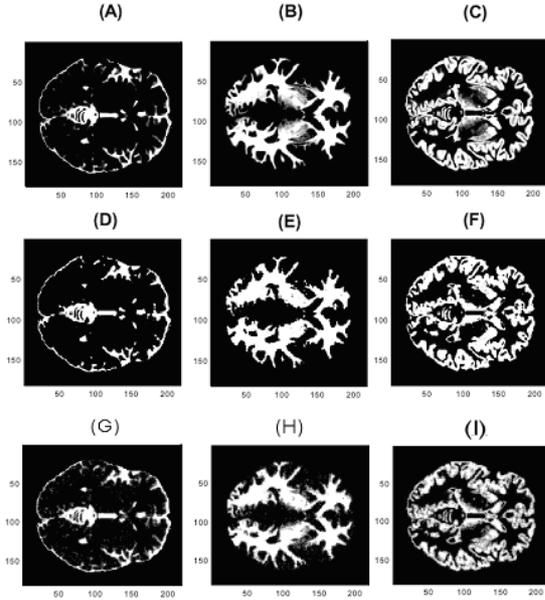


Figure 3: A) B) C): Fraction of CSF, WM and GM tissues in the digital phantom. D) E) F): Fraction of CSF, WM and GM classes obtained with the DM. G) H) I): Fractions obtained with the PVM.

Table I: Optimal values of the MAE and Evidence for the DM and PVM. In this case we find that the error is minimized and the Evidence maximized for $N_p=4$.

Model	MAE (Mean Absolute Error)	Log[Evidence]
DM	0.095	-3.75
PVM ($N_p=4$)	0.065	-3.65

where $p(j)$ is the fraction of tissue j and $p_{i,j}$ is the predicted value of tissue j in pixel i for the available models.

6.4 DM validation

Figures 3.A.B.C show t and classification p for the DM is shown in Figs. 3.D.E.F for tissues corresponding to GM, WM and CSF, respectively. Fig. 4 displays the behavior of MAE, the Evidence of the DM and the prediction of the most probable β according to (18). Note that the Evidence correctly predicts the behavior of the MAE.

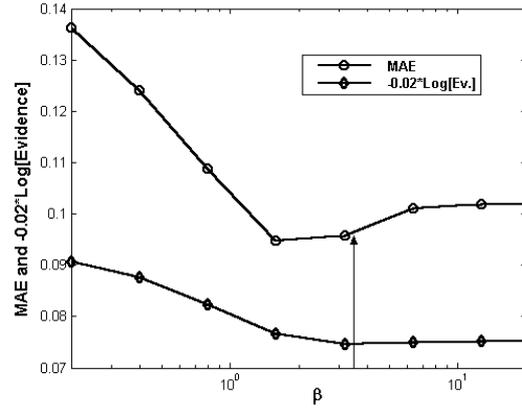


Figure 4: Behavior of MAE ("o"), the Evidence ("◇") of the DM as a function of β . The prediction of the most probable β according to (18) is shown with a vertical arrow. Note that the $-\text{Log}[\text{Evidence}]$ predicts the behavior of the MAE.

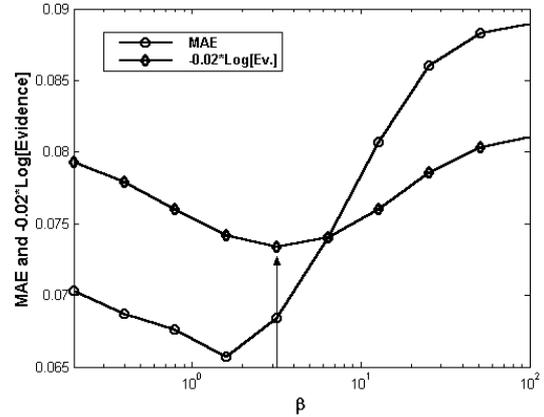


Figure 5: Comparison between MAE ("o") and $-\text{Log}[\text{Evidence}]$ ("◇") of the PVM as a function of β . The value of β that maximizes (24) is indicated with an arrow. The value of N_p the maximizes (24) is 4.

6.5 PVM validation

Figs. 3.G.H.I. show the classification p for the PVM for the same slice. Fig. 5 shows $-\text{Log}[\text{Evidence}]$ (24) and the MAE (28). Although the minimal values for the two curves do not match exactly, it is shown that a) the Maximum Evidence method provides a good estimation of β , and b) again, the Evidence proves that for high values of β , the DM is better than the PVM.

6.6 Model Comparison

The optimal values of $\text{Log}[\text{Evidence}]$ and MAE are listed in Table I. They correspond to each model

(classifications performed on the digital phantom). It shows that PVM has a lower error and a larger evidence than the DM.

7. Discussion and Conclusion

The analysis of brain MR images was treated in the present work as an unsupervised regression problem, adopting the Maximum Evidence criterion as the Bayesian method for optimization of parameters, hyperparameters and model comparison. As the numerical calculation for the Evidence is a quite burdensome task, an approximate approach was developed that significantly reduces calculation time.

Furthermore, it was shown that the Evidence is able to predict the behavior of the Mean Absolute Error. The estimations achieved for the pixel interaction factors worked out satisfactorily for the two models under consideration: Discrete (DM) and Partial Volume (PVM).

APPENDIX A

According to (1),(2) all the information relevant to the inference problem can be obtained from the joint distribution:

$$P(\mathbf{d}, \mathbf{w} | \gamma_i, H_i) = \frac{\exp[-E(\mathbf{w})]}{Z(\gamma_i)} \quad (\text{A.1})$$

$$Z(\gamma_i) = \int d\mathbf{d} \int d\mathbf{w} \cdot \exp[-E(\mathbf{w})]$$

where $E(\mathbf{w})$ is:

$$E(\mathbf{w}) = \sum_{i=1}^N e(w_i, w_{N(i)}) = \sum_{i=1}^N e_D(w_i) + \frac{\beta}{2} V(w_i, w_{N(i)}) \quad (\text{A.2})$$

with the approximation:

$$E(\mathbf{w}) \approx \sum_{i=1}^N e(w_i | \hat{w}_{N(i)}) = \sum_{i=1}^N e_D(w_i) + \frac{\beta}{2} V(w_i | \hat{w}_{N(i)}) \quad (\text{A.3})$$

Thus, the joint distribution is approximated for the whole image as the product of the distributions for each pixel:

$$P(\mathbf{d}, \mathbf{w} | \gamma_i, H_i) = \prod_{i=1}^N \frac{\exp[-e(w_i | \hat{w}_{N(i)})]}{Z_c^{(i)}(\gamma_i)} \quad (\text{A.4})$$

$$Z_c^{(i)}(\gamma_i) = \int dd_i \int dw_i \exp[-e(w_i | \hat{w}_{N(i)})]$$

8. References

- [1] J. Bezdek, L. Hall, and L. Clarke. "Review of MR image segmentation techniques using pattern recognition", *Med. Phys.*, vol. 20, no. 4, pp. 1033-1048, 1993.
- [2] P. Perona and J. Malik. "Scale space and edge detection using anisotropic diffusion", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, July 1990, pp. 629.
- [3] G. Gerig, O. Kübler, R. Kikinis and FA Jolesz. "Nonlinear anisotropic filtering of MRI data". *IEEE Trans. Med. Imag.*, vol. 11., no. 2, June 1992. pp. 221.
- [4] M. Schneider, P. Fieguth, W. Karl, and A. Willsky. "Multiscale methods for segmentation and reconstruction of signals and images". *IEEE Trans. Image Processing*, vol. 9, no. 3, March 2000, pp. 456-468.
- [5] J. Weickert, B. ter Haar Romeny, M. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering". *IEEE Trans. Image Processing*, vol. 7., no 3, March 1998, pp. 398-410.
- [6] Y. Zhang, M. Brady, and S. Smith. "Segmentation of brain images through a hidden Markov random field model and the expectation-maximization algorithm". *IEEE Trans. Med. Imag.*, vol. 20. no. 1. January 2001. Pag. 45.
- [7] W. Wells, R. Kikinis, W. Grimson and F. Jolesz. "Adaptative segmentation of MRI data". *IEEE Trans. Med. Imag.*, vol. 15, no. 4, August 1996, pp. 429.
- [8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, 1984, pp. 721-741.
- [9] K. Held, E. Rota Kops, B. Krause, W. Wells III, R. Kikinis, and H. Müller-Gärtner, "Markov random field segmentation of brain images". *IEEE Trans. Med. Imag.*, vol. 16., no. 6, December 1997. pp. 879.
- [10] D. MacKay, "Bayesian interpolation". *Neural Computation*, vol. 4, no. 3, 1992. pp. 415-447.
- [11] R. Neal, Bayesian training of backpropagation networks by the Hybrid Monte Carlo method. Technical Report CRG-TR-92-1, Dept. of Computer Science, University of Toronto, (1992).
- [12] D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes and A. Evans. "Design and construction of a realistic digital brain phantom", *IEEE Trans. Med. Imag.*, vol. 17, no. 3, June 1998, pp. 463-468.
- [13] J. Besag. "On the statistical analysis of dirty pictures". *J. Royal Statistical Soc. B*, vol. 48, no. 3, pp. 259-302, 1986.
- [14] M. Brummer, R. Mersereau, R. Eisner, R. Lewine, "Automatic detection of brain contours in MRI data sets", *IEEE Trans Med Imag*, vol. 12., no 2., June 1993, pp. 153.
- [15] S. Ruan, C. Jaggi, J. Xue, J. Fadili and D. Bloyet, "Brain tissue classification of magnetic resonance images using partial volume modeling", *IEEE Trans. Med. Imag.*, vol. 19. no. 12, December 2000, pp. 1179.
- [16] K. Van Leemput, F. Maes, D. Vandermeulen and P. Suetens, "A unifying framework for partial volume segmentation of brain MR images", *IEEE Trans. Med. Imag.*, vol. 22. no. 1, January 2003, pp. 105.